

Phân Tích Dữ Liệu Thực Tế với Python

Bài 9.1: Các Loại Biểu Đồ Nâng Cao



Quang-Khai Tran, Ph.D
CyberLab, 04/2023



(Ảnh: Internet)

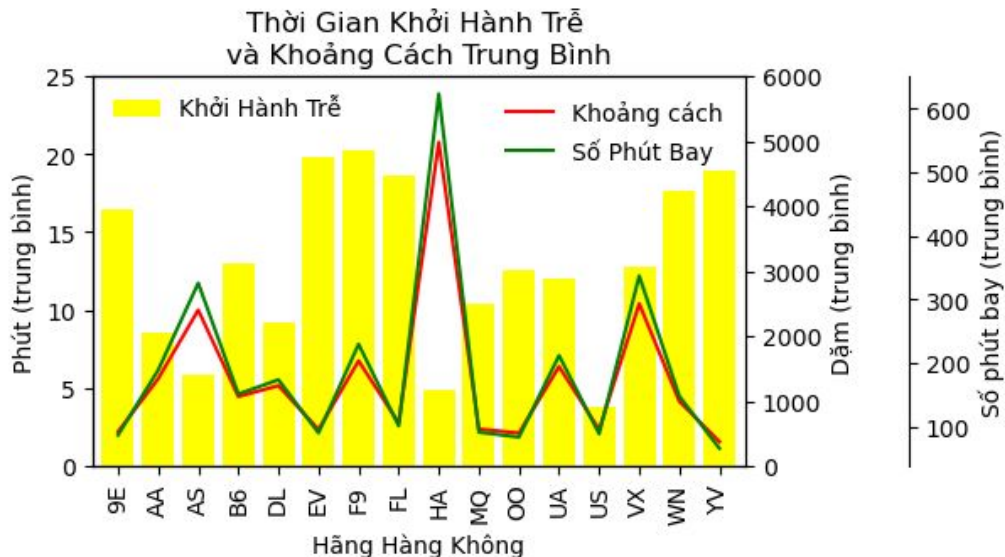
Nội dung



1. Biểu đồ với thông số khác loại
2. Biểu đồ bubble
3. Biểu đồ box
4. Biểu đồ donut/nested-donut
5. Biểu đồ radar
6. Biểu đồ nhiệt
7. Thảo Luận & Hỏi đáp

- ❖ Matplotlib hỗ trợ vẽ nhiều thông số có scale khác nhau trên cùng một biểu đồ
- ❖ Để tìm kiếm: **matplotlib + plotting with different scales**
- ❖ Tham khảo nâng cao:

https://matplotlib.org/devdocs/gallery/subplots_axes_and_figures/two_scales.html



❖ Cách thực hiện:

- Để thêm một scale, tạo một "song sinh" với biểu đồ đã vẽ
- Share trục x, khác trục y: `twinx()`
- Share trục y, khác trục x: `twiny()`

```
fig, ax1 = plt.subplots()  
ax2 = ax1.twinx()  
ax1.plot(...)  
ax2.plot(...)
```

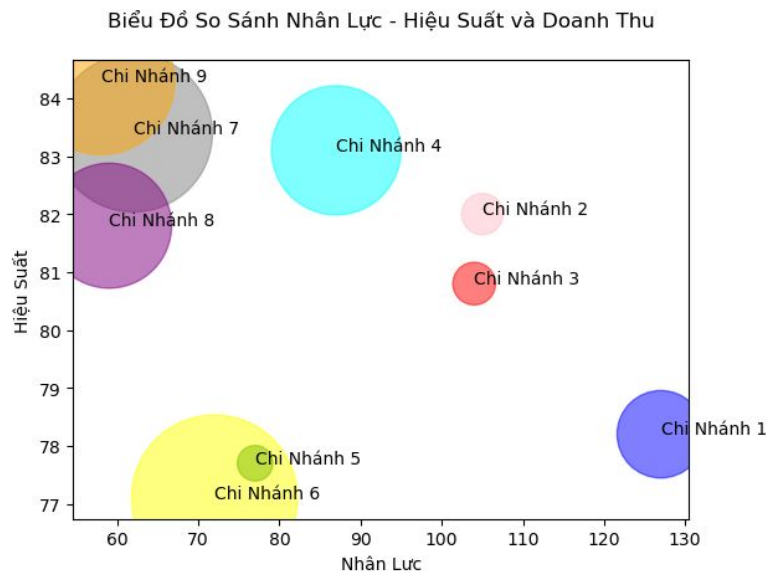
❖ Cách thực hiện:

- Trong trường hợp có 3 thông số, để hiển thị riêng một trục tung nữa cần set vị trí của spines

```
fig, ax1 = plt.subplots()
ax2 = ax1.twinx()
ax1.plot(...)
ax2.plot(...)

ax3 = ax1.twinx()
ax3.spines['right'].set_position(('axes', 1.25))
```

- ❖ Biểu đồ bubble (dạng bọt) là loại biểu đồ scatter với kích thước điểm biến đổi
- ❖ Kích thước lớn tương ứng với "cường độ" cao hơn của điểm dữ liệu



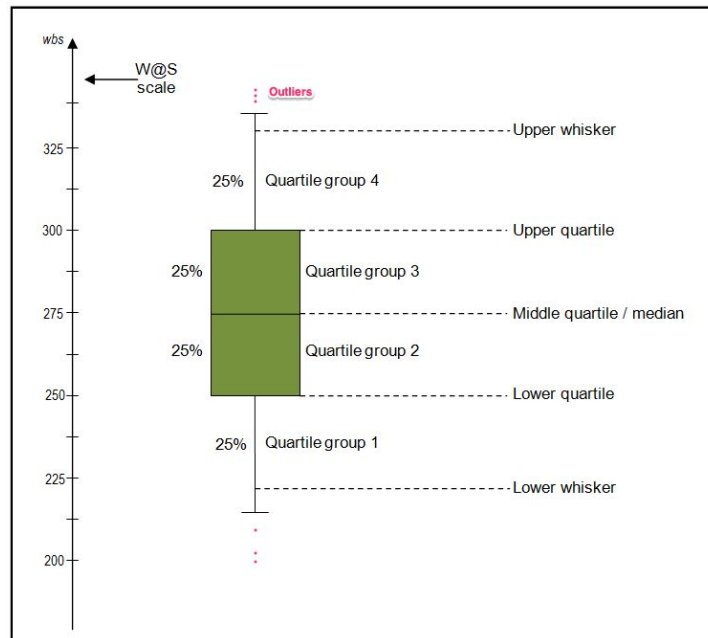
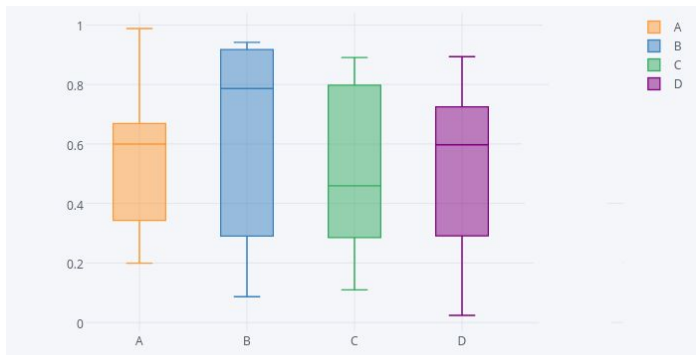
❖ Cách thực hiện:

- Set giá trị về kích thước cho từng điểm dữ liệu trong hàm vẽ scatter
- Có thể thêm giá trị về màu sắc trong trường hợp có ít điểm dữ liệu

```
plt.scatter(x=...,  
            y=...,  
            size/s=...,  
            color/c=...)
```

❖ Biểu đồ box (còn gọi là whisper-plot) dùng để tóm tắt các thông tin của một trường dữ liệu, bao gồm:

- minimum,
- first quartile,
- median,
- third quartile
- maximum



❖ Cách thực hiện:

- Với matplotlib

(Lưu ý: để set màu của box, cần set patch_artist=True)

- Với pandas

```
# Matplotlib
plt.boxplot(data, patch_artist=True)

# Pandas
df['list-các-cột'].plot.box()
df['list-các-cột'].boxplot()
```

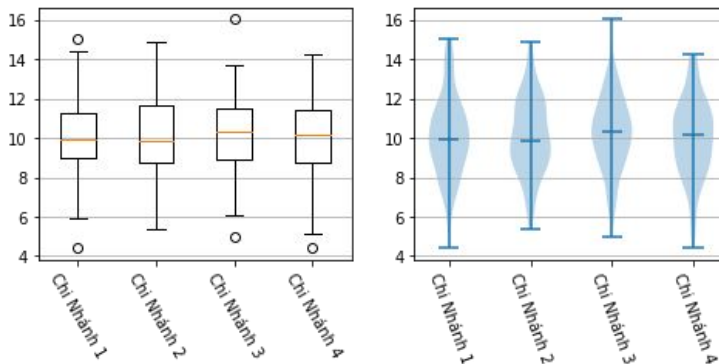
❖ Tham khảo thêm:

https://matplotlib.org/devdocs/gallery/statistics/boxplot_color.html

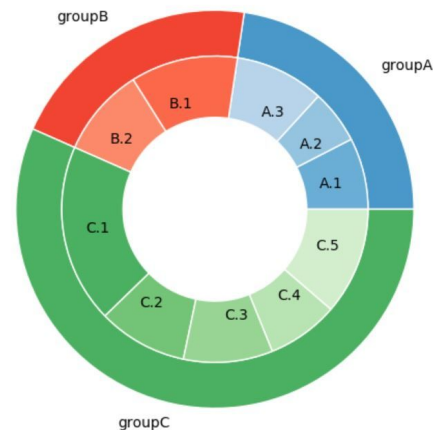
❖ Tương tự như boxplot là **violinplot**

```
# Matplotlib
ax2.violinplot(arr, showmeans=False,
               showmedians=True, showextrema=True,)

# Lưu ý: phải set vị trí của xticks
ax2.set_xticks([1,2,3,4])
ax2.set_xticklabels(labels, rotation=-70,
                  ha='center')
```



- ❖ Là dạng biểu đồ tròn có hình dạng như bánh donut
- ❖ Nested-donut: biểu đồ donut lồng ghép với nhau, hoặc với biểu đồ tròn



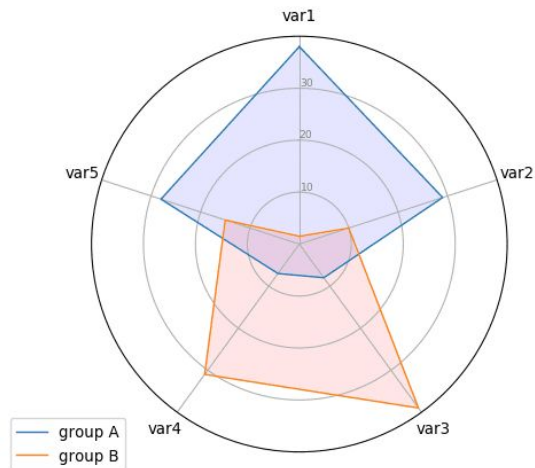
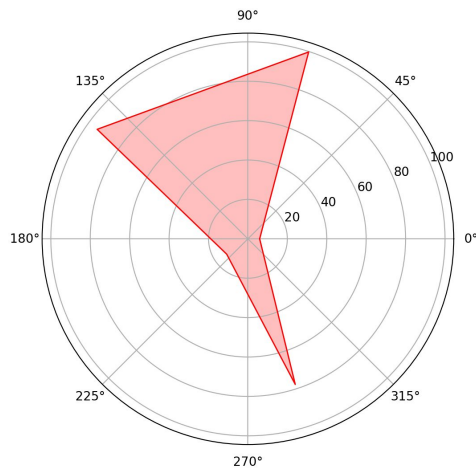
❖ Cách thực hiện:

- Cách 1: vẽ biểu đồ pie như thông thường và vẽ 1 hình tròn vào giữa
- Cách 2: trong hàm vẽ plt.pie, set giá trị `wedgeprops=dict(width=W)`

```
ax.pie(data, wedgeprops=dict(width=0.5))
```



- ❖ Còn gọi là biểu đồ "mạng nhện" (spider) hoặc biểu đồ "sao" (star)
- ❖ Biểu đồ radar giúp so sánh cường độ của
 - Các thuộc tính của một thông số
 - Các thuộc tính giống nhau của nhiều thông số



❖ Tham khảo:

https://matplotlib.org/stable/gallery/specialty_plots/radar_chart.html

❖ Có nhiều cách để tạo biểu đồ radar

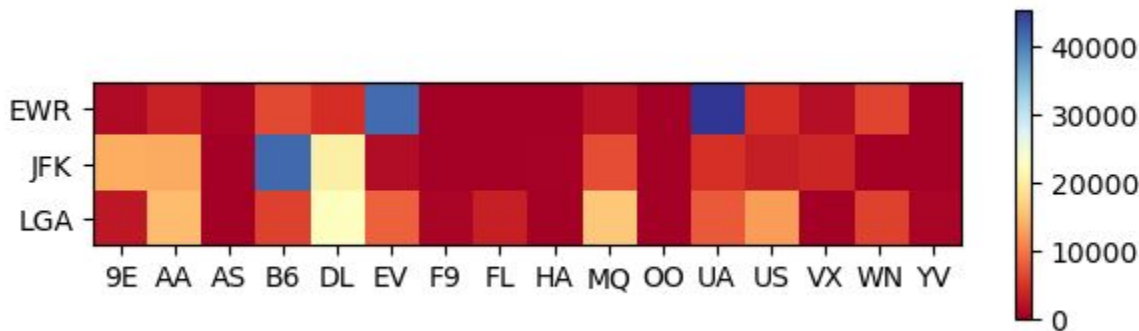
❖ Cách đơn giản:

- Dùng tham số `polar=True` hoặc `projection='polar'` khi gọi hàm `add_subplot()`
- Mục đích: biến đồ thị trở thành dạng có gốc tọa độ là tâm hình tròn
- Khi đó: giá trị x là góc của toa độ, giá trị y là khoảng cách từ tâm

```
fig = plt.figure(figsize=(4,4))
ax = fig.add_subplot(polar=True)
ax.plot([3.14/2, 3.14, 3.14*1.75, 3.14/2], [4, 6, 5, 4])
```

- ❖ Heatmap: hiển thị cường độ một bảng dữ liệu
- ❖ Có nhiều cách thực hiện. Ví dụ:
 - Dùng hàm hiển thị hình ảnh của matplotlib: `imshow()`
 - Hiển thị Pandas dataframe theo dạng có trang trí.

```
im = ax.imshow(numpy_arr, cmap = "RdYlBu")  
plt.colorbar(im)
```



THANK YOU!

